ChromaSig: a probabilistic approach to finding common chromatin signatures in the human genome.

Journal: PLoS Comput Biol

Publication Year: 2008

Authors: Gary Hon, Bing Ren, Wei Wang

PubMed link: 18927605

Funding Grants: Mapping the transcriptional regulatory elements in the genome of hESC

Public Summary:
The DNA in eukaryotes is packaged by histones. Interestingly, histones can be marked by a variety of posttranslational modifications, and it has been hypothesized that distinct combinations of histone modifications mark at distinct functional regions of the genome. The study of histone modifications has been aided by the development of high-throughput techniques to map a wide assortment of histone modifications on a global scale. However, because much of our current understanding of the human genome is concentrated on promoters, most studies have only examined histone modifications at these well-defined sites, ignoring the vast majority of the genome. To aid in the discovery of functional elements outside of these well-annotated loci, we develop an unbiased method that searches for commonly occurring histone modification patterns on a global scale without using any annotation information. This method recovers known patterns associated with transcriptional enhancers and promoters. Supporting the histone code hypothesis, we discover that the different functional activities of enhancers are closely associated with the presence of different histone modification patterns. We also discover several novel patterns that likely contain other potential regulatory elements. As the availability of large-scale histone modification data increases, the ability of methods such as the one presented here to concisely describe commonly occurring chromatin signatures, thereby abstracting away irrelevant or redundant data, will become increasingly more critical.

Scientific Abstract:
Computational methods to identify functional genomic elements using genetic information have been very successful in determining gene structure and in identifying a handful of cis-regulatory elements. But the vast majority of regulatory elements have yet to be discovered, and it has become increasingly apparent that their discovery will not come from using genetic information alone. Recently, high-throughput technologies have enabled the creation of information-rich epigenetic maps, most notably for histone modifications. However, tools that search for functional elements using this epigenetic information have been lacking. Here, we describe an unsupervised learning method called ChromaSig to find, in an unbiased fashion, commonly occurring chromatin signatures in both tiling microarray and sequencing data. Applying this algorithm to nine chromatin marks across a 1% sampling of the human genome in HeLa cells, we recover eight clusters of distinct chromatin signatures, five of which correspond to known patterns associated with transcriptional promoters and enhancers. Interestingly, we observe that the distinct chromatin signatures found at enhancers mark distinct functional classes of enhancers in terms of transcription factor and coactivator binding. In addition, we identify three clusters of novel chromatin signatures that contain evolutionarily conserved sequences and potential cis-regulatory elements. Applying ChromaSig to a panel of 21 chromatin marks mapped genomewide by ChIP-Seq reveals 16 classes of genomic elements marked by distinct chromatin signatures. Interestingly, four classes containing enrichment for repressive histone modifications appear to be locally heterochromatic sites and are enriched in quickly evolving regions of the genome. The utility of this approach in uncovering novel, functionally significant genomic elements will aid future efforts of genome annotation via chromatin modifications.